

Evaluation of space-time point process models using super-thinning

Robert Alan Clements¹, Frederic Paik Schoenberg¹, and Alejandro
Veen²

¹UCLA Department of Statistics, 8125 Math Sciences Building, Los Angeles, CA 90095-1554

²IBM, T.J. Watson Research Center, Yorktown Heights, NY 10598

Abstract

Rescaling, thinning and superposition are useful methods for the residual analysis of spatial-temporal point processes. These techniques involve transforming the original point process into a new process that should be a homogeneous Poisson process if and only if the fitted model is correct, so that one may inspect the residual process for homogeneity using standard tests for homogeneity as a means of assessing the goodness-of-fit of the model. Unfortunately, tests of homogeneity performed on residuals based on these three residual methods tend to have low power when the modeled conditional intensity of the original process is volatile. For such purposes, we propose the method of super-thinning, which combines thinned residuals and superposition. This technique involves the use of a tuning parameter, k , which controls how much thinning and superposition are performed to homogenize the process. The method is applied to the assessment of a parametric space-time point process model for the origin times and epicentral locations of recent major California earthquakes.

1 Introduction

Residual analysis methods for spatial point process models, such as thinning, rescaling, and superposition, involve transforming the point process using a model for the conditional intensity λ and then inspecting the uniformity of the result. The difficult problem of evaluating the agreement between a possibly complex spatial-temporal point process model and data thus boils down to the apparently simpler problem of assessing the homogeneity of the residual point process, a task for which many standard tests are available.

The method of randomly thinning a point process is based on the works of Lewis and Shedler (1979); Ogata (1981), who used random thinning as a means of simulating a spatial-temporal point process, and was used for evaluating models for earthquake occurrences in Schoenberg (2003). The points that remain after thinning, called thinned residual points, form a homogeneous Poisson process if and only if the model for λ used to perform the thinning is correct. A completely opposite approach was proposed by Brémaud (1981), who suggested superposing a simulated point process onto an observed point process realization so as to yield a homogeneous Poisson process. Meyer (1971) investigated rescaling point processes according to their conditional intensities, moving each point to a new time (or location), creating a transformed space in which the rescaled points are homogeneous Poisson of unit rate. This method was used in Ogata (1988) to assess models for earthquake times and magnitudes and extended in Merzbach and Nualart (1986), Nair (1990), Schoenberg (1999), and Vere-Jones and Schoenberg (2004) to the spatial-temporal case.

Unfortunately, tests based on the residuals formed from each of these methods tend

to have low power when the model $\hat{\lambda}$ for the conditional intensity of the original point process is volatile. Thinning a point process will lead to very few points remaining if the infimum of $\hat{\lambda}$ over the observed space is small (see Schoenberg (2003) for an example), while in superposition, the simulated points, which are by construction approximately homogeneous, will form the vast majority of residual points if the supremum of $\hat{\lambda}$ is large. Rescaling may result in a transformed space that is difficult to inspect if $\hat{\lambda}$ varies widely over the observation region, and in such cases standard tests of homogeneity such as Ripley's K-function (Ripley (1981)) may be dominated by boundary effects, as in Schoenberg (2003).

A more powerful approach is super-thinning, which is a combination of thinning observed points and superimposing simulated points, leading to a homogeneous residual point process if and only if the estimate of the conditional intensity λ of the original point process is correct. With super-thinning, the relative amount of thinning and superposition can be controlled through the choice of a tuning parameter, k . With a prudent choice of k , the method can be substantially more powerful than either thinning or superposition alone.

In Section 2 we briefly review thinning and superposition and discuss their limitations. The proposed method of super-thinning is described in Section 3 and the proof that the resulting residual process is Poisson if and only if the estimated conditional intensity is correct almost everywhere is provided, along with some criteria for choosing the value of the tuning parameter, k . We illustrate the three methods with several simulated examples, and apply super-thinning to test the goodness-of-fit of a California earthquake forecast model in Section 4.

2 Thinning and Superposition

2.1 Preliminaries

Assume throughout that N is a σ -finite spatial-temporal point process on a compact set $\mathcal{S} \subset \mathbb{R}^d$, adapted to a filtration \mathcal{F} equipped with probability measure P . Let $\{(\mathbf{x}_i, t_i); i = 1, 2, \dots, n\}$ denote the collection of observed points of N . Let μ denote Lebesgue measure on \mathcal{S} , and let $\pi = \mu \times P$ denote the product measure on $P \times \mathcal{S}$. The *compensator* A of N is the unique non-negative \mathcal{F} -predictable process such that $N - A$ is an \mathcal{F} -martingale, and the Radon-Nikodym derivative of A , if it exists, is the \mathcal{F} -conditional intensity of N . We assume in what follows that N has \mathcal{F} -conditional intensity $\lambda(\mathbf{x}, t) < \infty$.

Suppose throughout that N has simple ground process, i.e. that N has at most one point at any *time*, with probability one. Let $\hat{\lambda}(\mathbf{x}, t)$ be an \mathcal{F} -predictable estimate (either parametric or nonparametric) of the conditional intensity $\lambda(\mathbf{x}, t)$. For brevity, we refer to \mathcal{F} -predictability in what follows simply as predictability. Note that a variety of different types of conditional intensity corresponding to different types of conditioning may be available. We assume in what follows that $\mathcal{F}(\mathbf{x}, t)$ contains information on N for all previous *times*, i.e. $\mathcal{F}(\mathbf{x}_1, t) \subseteq \mathcal{F}(\mathbf{x}_2, u)$ for $t < u$. For a review of basic definitions related to spatial-temporal point processes and conditional intensities, see Daley and Vere-Jones (2003), Brillinger and Guttorp (2002), or Schoenberg (1999).

2.2 Thinned residuals

In residual thinning, each observed point is retained independently with probability $b/\hat{\lambda}(\mathbf{x}_i, t_i)$, where $b = \inf_{(\mathbf{x}, t) \in \mathcal{S}} \{\hat{\lambda}(\mathbf{x}, t)\}$. The points that remain are called *thinned*

residual points, and if the estimate $\hat{\lambda}(\mathbf{x}, t) = \lambda(\mathbf{x}, t)$ almost everywhere, then the residual process, Z , will be homogeneous Poisson with rate b (Schoenberg (2003)). To detect clustering or inhibition in the residual point pattern, two widely used statistics are Ripley’s K-function (see Ripley (1981)), and the variance stabilized version of the K-function called the L-function (see Besag (1977)). In practice, one may generate several realizations of thinned residuals and analyze each of them to get an overall assessment of the goodness-of-fit, as in Schoenberg (2003).

The power of thinned residuals may suffer in part due to the variability and lack of independence in the thinned residual points and especially due to the loss of information when removing observed points. Indeed, if b is small, as is often the case when modeling spatially inhomogeneous phenomena such as earthquakes or wildfires, then thinning may result in very few residual points, so that tests, both formal and informal, will have little power to detect inhomogeneity. One may increase the number of residual points retained by instead keeping each point with probability $c/\hat{\lambda}(\mathbf{x}_i, t_i)$ as in (Peng (2003)), where $c > b$ is some constant selected by the user. Provided that c is small relative to the mean of λ , the resulting point process will be approximately homogeneous Poisson if the model for the conditional intensity λ is correct. However, if c is large, the thinned residuals may exhibit substantial inhomogeneity even if $\hat{\lambda} = \lambda$, and distinct thinnings will be highly correlated (Schoenberg (2003)).

2.3 Superposition

Residual superposition involves transforming the point process N into a homogeneous Poisson process by simulating points rather than removing points. A residual point process is created by simulating a Poisson process with intensity $d - \hat{\lambda}(\mathbf{x}, t)$, where $d = \sup_{(\mathbf{x}, t) \in \mathcal{S}} \{\hat{\lambda}(\mathbf{x}, t)\}$, i.e. a Cox process directed by $d - \hat{\lambda}(\mathbf{x}, t)$ if $\hat{\lambda}$ is random. The simulated process is then superposed onto N , creating a homogeneous Poisson point

process of rate d iff. $\hat{\lambda} = \lambda$ almost everywhere. Any significant clustering, inhibition, or inhomogeneity in the superposed residuals indicates a lack of fit of the model $\hat{\lambda}$, and as with thinning, several realizations of superposed residuals may be generated and assessed for uniformity.

Residual analysis using superposition has limited power when the supremum of $\hat{\lambda}$ is much larger than the mean of $\hat{\lambda}$ over the observation region \mathcal{S} , as is often the case with models of highly inhomogeneous or clustered phenomena. In such cases, the simulated points are by construction Poisson and approximately homogeneous in regions where $d \gg \hat{\lambda}$, and since the number of such simulated points is large, their inclusion in the residuals tends to overwhelm any information provided by N in standard tests of homogeneity, resulting in little power to detect inhomogeneity in the residuals as a whole.

3 Super-thinning

We propose a hybrid approach, involving both thinning and superposition. Suppose that one desires to transform the point process N into a residual Poisson process with rate k , where $\inf\{\hat{\lambda}(\mathbf{x}_i, t_i)\} \leq k \leq \sup\{\hat{\lambda}(\mathbf{x}_i, t_i)\}$. One may first thin N , keeping each point (\mathbf{x}_i, t_i) independently with probability $\min\{k/\hat{\lambda}(\mathbf{x}_i, t_i), 1\}$ to obtain a thinned residual process Z_1 . Next, one simulates a Cox process Z_2 directed by $\max\{k - \hat{\lambda}(\mathbf{x}, t), 0\}$. That is, conditional on k and $\hat{\lambda}$, Z_2 is a Poisson process whose rate at location (\mathbf{x}, t) is $\max\{k - \hat{\lambda}(\mathbf{x}, t), 0\}$. Note that this second step can easily be performed by simulating a homogeneous Poisson process with rate k and independently keeping each simulated point $(\tilde{\mathbf{x}}_j, \tilde{t}_j)$ with probability $\max\{(k - \hat{\lambda}(\tilde{\mathbf{x}}_j, \tilde{t}_j))/k, 0\}$. The points of the residual point process $Z = Z_1 + Z_2$, obtained by superposing the thinned residuals and the simulated Poisson process, are called *super-thinned residual points*. Because Z is homogeneous Poisson with rate k if and only if $\hat{\lambda} = \lambda$ almost everywhere,

as shown below, one may inspect the points in Z for uniformity as a way of assessing the goodness-of-fit of the estimate $\hat{\lambda}$.

The super-thinned residual process, Z , may be formally defined as follows. Let W_t be a Uniform(0,1) white noise process adapted to \mathcal{F} and independent of N . Let Q be a Poisson process with rate k adapted to \mathcal{F} and independent of N and W .

Let $\hat{\lambda} > 0$ denote an estimate of λ , and define $Z_1(\mathbf{x}, t) = N(\mathbf{x}, t)\mathbf{1}_{(W_t < k/\hat{\lambda}(\mathbf{x}, t))}$ and $Z_2(\mathbf{x}, t) = Q(\mathbf{x}, t)\mathbf{1}_{(W_t < (k-\hat{\lambda}(\mathbf{x}, t))/k)}$. The superposition $Z = Z_1 + Z_2$ defines the super-thinned residual process.

Theorem 3.1. Z is a Poisson process with rate k iff. $\hat{\lambda} = \lambda$ π -a.e.

Proof. Note that if we can show that Z has simple ground process and has compensator $k\mu$, then Z is a Poisson process with rate k by Proposition 4.2 of Nair (1990). Since N has simple ground process, and since Q is a Poisson process independent of N , it is trivially true that the superposition $N + Q$ has simple ground process, and therefore the same is true of Z .

We now show that if $\hat{\lambda} = \lambda$ a.e., then Z has compensator $k\mu$. Fix t and \mathbf{x} and let F be any event in $\mathcal{F}_{\mathbf{x}, t}$. Let C denote a set of the form $(t, u) \times X$, where X is a

measurable subset of the spatial domain. Observe that, since $0 \leq W(y) \leq 1$,

$$\begin{aligned}
E[Z(C)|F] &= E \left[\int_C \mathbf{1}_{(\hat{\lambda}(y) \geq k)} \mathbf{1}_{(W(y) < \frac{k}{\hat{\lambda}(y)})} dN(y) | F \right] \\
&\quad + E \left[\int_C \mathbf{1}_{(\hat{\lambda}(y) < k)} dN(y) | F \right] \\
&\quad + E \left[\int_C \mathbf{1}_{(\hat{\lambda}(y) < k)} \mathbf{1}_{(W(y) < \frac{k - \hat{\lambda}(y)}{k})} dQ(y) | F \right] \\
&= E \left[\int_C E \left[\mathbf{1}_{(\hat{\lambda}(y) \geq k)} \mathbf{1}_{(W(y) < \frac{k}{\hat{\lambda}(y)})} | F_y \right] dN(y) | F \right] \\
&\quad + E \left[\int_C \mathbf{1}_{(\hat{\lambda}(y) < k)} dN(y) | F \right] \\
&\quad + E \left[\int_C E \left[\mathbf{1}_{(\hat{\lambda}(y) < k)} \mathbf{1}_{(W(y) < \frac{k - \hat{\lambda}(y)}{k})} | F_y \right] dQ(y) | F \right] \\
&= E \left[\int_C \mathbf{1}_{(\hat{\lambda}(y) \geq k)} \frac{k}{\hat{\lambda}(y)} dN(y) | F \right] \\
&\quad + E \left[\int_C \mathbf{1}_{(\hat{\lambda}(y) < k)} dN(y) | F \right] \\
&\quad + E \left[\int_C \mathbf{1}_{(\hat{\lambda}(y) < k)} \frac{k - \hat{\lambda}(y)}{k} dQ(y) | F \right]
\end{aligned}$$

Thus by the martingale property (see e.g. equation (1) of Nair (1990)),

$$\begin{aligned}
E[Z(C)|F] &= E \left[\int_C \mathbf{1}_{(\hat{\lambda}(y) \geq k)} \frac{k}{\hat{\lambda}(y)} \lambda(y) d\mu | F \right] \\
&+ E \left[\int_C \mathbf{1}_{(\hat{\lambda}(y) < k)} \lambda(y) d\mu | F \right] \\
&+ E \left[\int_C \mathbf{1}_{(\hat{\lambda}(y) < k)} \frac{(k - \hat{\lambda}(y))}{k} (k) d\mu | F \right].
\end{aligned} \tag{1}$$

If $\hat{\lambda} = \lambda$ π -a.e., then equation (1) reduces immediately to

$$E[Z(C)|F] = (k)E \left[\int_C \mathbf{1}_{(\lambda(y) \geq k)} + \int_C \mathbf{1}_{(\lambda(y) < k)} d\mu | F \right] = k\mu(C),$$

so that $k\mu$ is the compensator of Z and thus Z is a Poisson process with rate k .

The converse may be shown as in the proof of Theorem 3.2 of Schoenberg (1999). Indeed, if $\hat{\lambda}$ is not equal to λ π -a.e., then at least one of the four sets $B_1 = \{\hat{\lambda} > \lambda\} \cup \{\hat{\lambda} \geq k\}$, $B_2 = \{\hat{\lambda} < \lambda\} \cup \{\hat{\lambda} \geq k\}$, $B_3 = \{\hat{\lambda} > \lambda\} \cup \{\hat{\lambda} < k\}$, or $B_4 = \{\hat{\lambda} < \lambda\} \cup \{\hat{\lambda} < k\}$ must have π -measure greater than zero. Without loss of generality, suppose that $\pi(B_1) > 0$. Then the indicator 1_{B_1} is predictable, since $\hat{\lambda}$ and λ are predictable, so as in (1),

$$E \int 1_{B_1} dZ = E \int 1_{B_1} \frac{k}{\hat{\lambda}(y)} \lambda(y) d\mu < kE \int 1_{B_1} d\mu,$$

so Z is not a Poisson process with rate k , since for a Poisson process with rate k , one would have equality in the above relation by the martingale property. \square

3.1 The tuning parameter

Note that in super-thinning, the tuning parameter k allows the user to control the rate of thinning and superposition. As a result, superthinning is potentially much more powerful than either thinning or superposition alone. The parameter k should be chosen in a way that optimizes the power of formal tests of homogeneity of the residuals, and is an ongoing problem that requires further study. Here, we introduce two practical methods for choosing k .

One suggested approach is to super-thin a point process by thinning out and superposing as few points as possible in order to retain as much of the original data as possible. This suggests choosing the value of k minimizing the sum of the absolute deviations of the estimated conditional intensity from k , i.e. solving

$$\operatorname{argmin}_k \int \int_{\mathcal{S}} \int_t |\hat{\lambda}(\mathbf{x}, t) - k| dt dx dy,$$

i.e. letting k equal the median of $\hat{\lambda}$. Note that in some cases a unique median might not exist, however.

Alternatively, one may choose the value of k minimizing the sum of the squared deviations of the estimated conditional intensity from k , letting k equal the mean of $\hat{\lambda}$,

$$k = \frac{1}{|\mathcal{S}|} \int \int_{\mathcal{S}} \int_t \hat{\lambda}(\mathbf{x}, t) dt dx dy,$$

where $|\mathcal{S}|$ is the volume of the observation region. In this case, the expected number of points in the resulting residual process is equal to the number of points as N . Depending on one's choice of test statistic performed on the residual process, it may be possible to choose k so that the resulting residual test has optimal power.

4 Examples

4.1 Inhomogeneous Poisson processes

Figure 1 illustrates super-thinning a point process on a space-time observation region $\mathcal{S} = [0, 2] \times [0, 2] \times [0, 1]$ divided into four equal bins, so that each bin is $[1 \times 1 \times 1]$, with constant conditional intensity in each bin, where

$$\lambda(\mathbf{x}, t) = \begin{cases} 80 & \text{for } 1 \leq y \leq 2 \text{ and } 0 \leq x \leq 1; \\ 20 & \text{else.} \end{cases}$$

Figure 1 shows the super-thinning of the simulated process N . The clustering in N is confirmed using a nonparametric estimate of the variance stabilized version of Ripley's K-function called the L-function, estimated by $\hat{L}(r) = \sqrt{\hat{K}(r)/\pi}$, where r represents distance. Figure 1(b) shows the estimated centered L-function, $\hat{L}(r) - r$, and 95% confidence bounds based on 1000 simulations of homogeneous Poisson processes with the same rate as N . For a homogeneous Poisson process, $L(r) - r = 0$, so departures from 0 indicate inhomogeneity. The super-thinned residuals in Figure 1(c), using the median $k = 20$ as the tuning parameter, are evidently homogeneous, as seen in Figure 1(d) by the estimated centered L-function, which is entirely within the 95% bounds.

The next example uses the same observation region, but with

$$\lambda(\mathbf{x}, t) = \begin{cases} 20 & \text{for } 1 \leq y \leq 2 \text{ and } 0 \leq x \leq 1; \\ 80 & \text{else.} \end{cases}$$

As seen in Figure 2(a), N clearly requires more superposition than thinning, and using the median $k = 80$, points are only superposed in the upper-left quadrant. Again, the super-thinned residuals are homogeneous, as shown in Figures 2(c) and (d).

Figures 3 and 4 are useful for comparing super-thinning to either thinning or superposition individually. In this illustration, $\mathcal{S} = [0, 1] \times [0, 1] \times [0, 1]$ with conditional intensity $\lambda(\mathbf{x}, t) = 3000 \exp(3x - 4y)$. Figure 3(a) is a realization of N , and points are highly clustered in the lower-left region, while being very sparse elsewhere. It is obvious that thinning or superposition alone would have low power in this example. Figure 3(c) and (d) show a typical result of thinned residuals and superposition, respectively, both of which reveal the primary weaknesses of both methods. Setting the tuning parameter to the mean $k = 233.2023$, one may perform super-thinning on this process, and Figure 4(a) shows the resulting residual process Z , whose homogeneity is confirmed by the centered L-function in Figure 4(b).

References

- J.E. Besag. Comments on “modelling spatial patterns” by b.d. ripley. *Journal of the Royal Statistical Society*, B39(2):193–195, 1977.
- P. Brémaud. *Point Processes and Queues: Martingale Dynamics*. SpringerVerlag, New York, 1981.
- F.P. Schoenberg D.R. Brillinger and P.M. Guttorp. Point processes, spatial-temporal. In Abdel El-Shaarawi and Walter Piegorsch, editors, *Encyclopedia of Environmental Metrics*, pages 1573–1577. Wiley, New York, 2002.
- D. Daley and D. Vere-Jones. *An Introduction to the Theory of Point Processes, volume 1*. SpringerVerlag, New York, 2003.
- P. Lewis and G. Shedler. Simulation of nonhomogeneous poisson processes by thinning. *Naval Research Logistics Quarterly*, 26:403–413, 1979.
- E. Merzbach and D. Nualart. A characterization of the spatial poisson process and changing time. *Annals of Probability*, 14:1380–1390, 1986.
- P. Meyer. Demonstration simplifiée d’un théorème de knight. *Séminaire de Probabilités V*, 191:191–195, 1971.
- M. Nair. Random space change for multiparameter point processes. *Annals of Probability*, 18:1222–1231, 1990.
- Y. Ogata. On lewis’ simulation method for point processes. *IEEE Transactions on Information Theory*, IT-27:23–31, 1981.
- Y. Ogata. Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical Association*, 83:9–27, 1988.

- R.D. Peng. Multi-dimensional point process models in r. *Journal of Statistical Software*, 8(16):1–27, 2003.
- B. Ripley. *Spatial Statistics*. Wiley, New York, 1981.
- F. Schoenberg. Transforming spatial point processes into poisson processes. *Stochastic Processes and their Applications*, 81:155–164, 1999.
- F.P. Schoenberg. Multi-dimensional residual analysis of point process models for earthquake occurrences. *Journal of the American Statistical Association*, 98:789–795, 2003.
- D. Vere-Jones and F.P. Schoenberg. Rescaling marked point processes. *Australian and New Zealand Journal of Statistics*, 46(1):133–143, 2004.

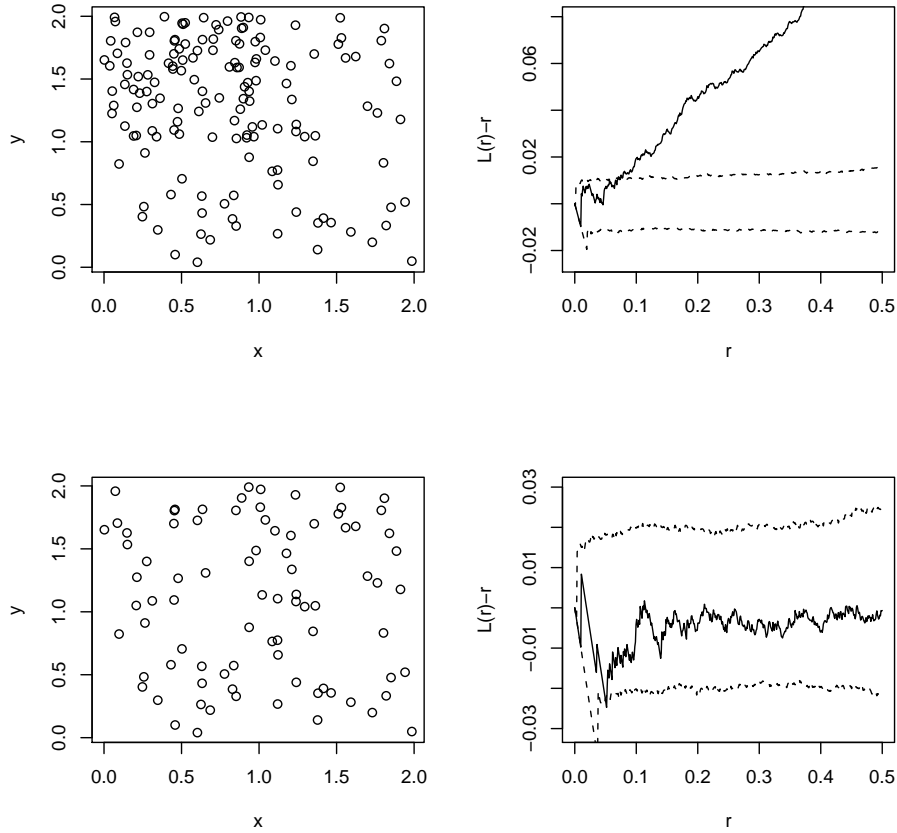


Figure 1: Super-thinning of a simulated inhomogeneous Poisson process with grid-based conditional intensity of $(80 \ 20 \ 20 \ 20)$ in upper-left, upper-right, lower-left and lower-right bin, respectively. Top-left panel (a): simulated inhomogeneous Poisson process, N . Top-right panel (b): estimated centered L-function for N with 95% bounds for a homogeneous Poisson process based on 1000 simulations of a homogeneous Poisson process with the same rate as N . Bottom-left panel (c): residual process, Z , obtained from super-thinning with $k = 20$. Bottom-right panel (d): estimated centered L-function for Z with 95% bounds for a homogeneous Poisson process based on 1000 simulations of a homogeneous Poisson process with the same rate as Z .

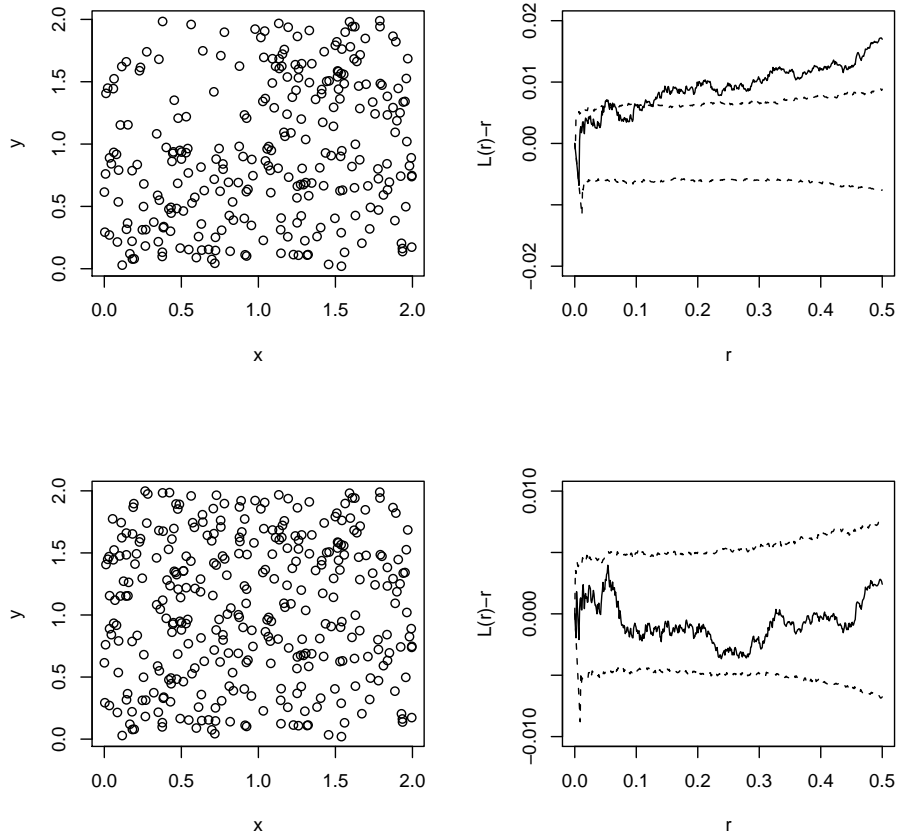


Figure 2: Super-thinning of a simulated inhomogeneous Poisson process with grid-based conditional intensity of $(20 \ 80 \ 80 \ 80)$ in upper-left, upper-right, lower-left and lower-right bin, respectively. Top-left panel (a): simulated inhomogeneous Poisson process, N . Top-right panel (b): estimated centered L-function for N with 95% bounds for a homogeneous Poisson process based on 1000 simulations of a homogeneous Poisson process with the same rate as N . Bottom-left panel (c): residual process, Z , obtained from super-thinning with $k = 80$. Bottom-right panel (d): estimated centered L-function for Z with 95% bounds for a homogeneous Poisson process based on 1000 simulations of a homogeneous Poisson process with the same rate as Z .

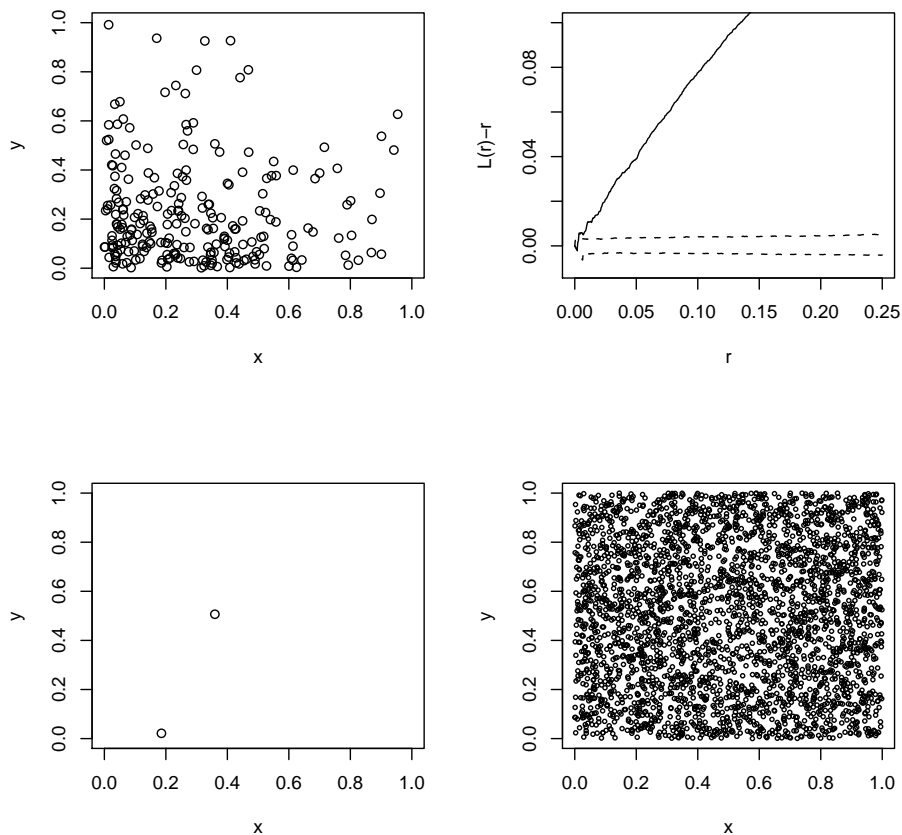


Figure 3: Thinned residuals and superposition of a simulated inhomogeneous Poisson process with conditional intensity function $\lambda(\mathbf{x}, t) = 3000e^{(-3x-4y)}$. Top-left panel (a): simulated inhomogeneous Poisson process, N . Top-right panel (b): estimated centered L-function for N with 95% bounds for a homogeneous Poisson process based on 1000 simulations of a homogeneous Poisson process with the same rate as N . Bottom-left panel (c): residual process, Z , obtained from standard thinned residuals. Bottom-right panel (d): residual process, Z obtained from standard superposition.

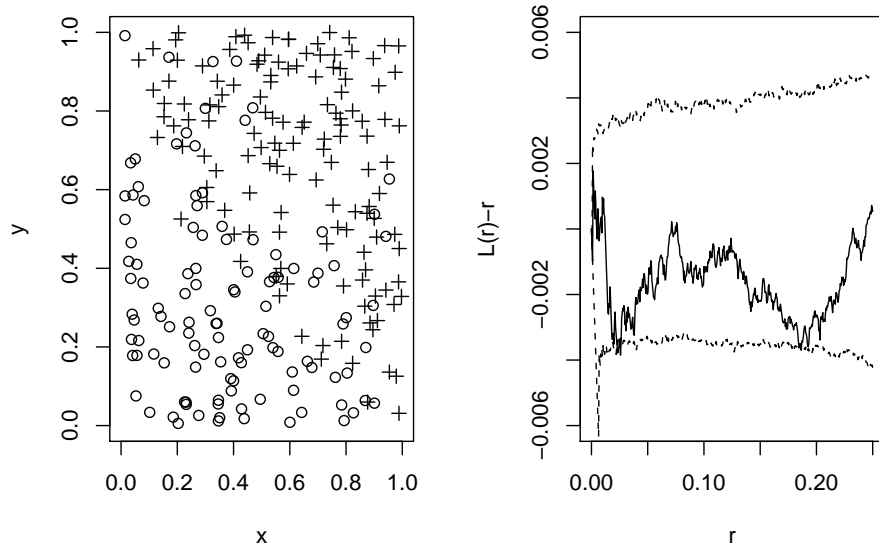


Figure 4: Super-thinning of a simulated inhomogeneous Poisson process with conditional intensity function $\lambda(\mathbf{x}, t) = 3000e^{(-3x-4y)}$. Left panel (a): residual process, Z obtained from super-thinning with $k = 233.2023$ (circles = observed points, plus signs = superposed points). Right panel (b): estimated centered L-function for Z with 95% bounds for a homogeneous Poisson process based on 1000 simulations of a homogeneous Poisson process with the same rate as Z .